

# 第1讲 大数据概述

2020年2月17日 10:36

信息化浪潮	发生时间	标志	解决问题	代表企业
第一次浪潮	1980年前后	个人计算机	信息处理	Intel、AMD、IBM、苹果、微软、联想 惠普、戴尔等
第二次浪潮	1995年前后	互联网	信息传输	雅虎、谷歌、阿里巴巴、百度、腾讯等
第三次浪潮	2010年前后	物联网、云计算和 大数据	信息爆炸	将涌现出一批新的市场标杆企业

## 1.2 大数据的概念和影响

在思维方式方面，大数据完全颠覆了传统的思维方式

全样而非抽样、效率而非精确、相关而非因果

## 1.3 大数据的应用

影视剧的投拍风险、谷歌预测流感

## 1.4 大数据关键技术

大数据技术的层次：

数据采集——> **数据存储与管理**——> **数据处理与分析**——> 数据隐私与安全

(这两层代表了很多非常核心的大数据技术)

两大核心技术：**分布式存储**、**分布式处理**

分布式数据库、分布式文件系统、分布式并行处理技术 (MapReduce)

### 典型的处理技术：

#### 批处理：

针对大规模数据的批量处理

1. MapReduce 是批处理计算模式的典型代表

2. Spark 高效的迭代计算

#### 流计算：

针对流数据的实时计算

雅虎: S4

流数据需要实时处理，给出实时响应，否则分析结果就会失去商业价值

#### 图计算：

针对大规模图结构数据的处理

Goole pregel

社交网络数据等

### **查询分析计算:**

大规模数据的存储管理和查询分析

代表产品: Google Dremel、Hive、Cassandra

## 1.5 大数据与云计算、物联网

1.5.1云计算的优势: 企业不需要自建IT基础设施, 只需租用云设备

公用云: 百度云等	私用云; 电信、移动	混合云:
-----------	------------	------

IaaS——基础设施及服务: 将基础设施(计算资源和存储)作为服务出租

PaaS——平台即服务: 在平台上开发在平台上部署

SaaS——典型案例: 云财务软件: 以服务的方式把财务软件系统卖给你, 不用本地安装

虚拟化: Hadoop、Linux系统、虚拟机、Windows系统、硬件、VPN虚拟网

多租户:

云计算数据中心: 数据中心包含大量刀片服务器

云计算运用: 教育云、医疗云、政府云